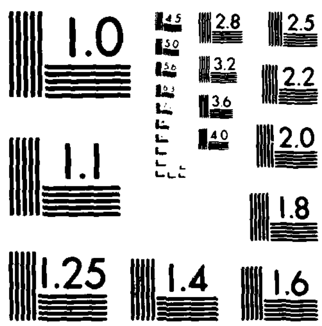
AD-A093 710 FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/G 12/1
SELECTING ALL TREATMENTS BETTER THAN A CONTROL USING EXISTING T--ETC(U)
OCT 80 R L BERGER DAAG29-79-C-0158
UNCLASSIFIED FSU-STATISTICS-M660 ARO-16713.13-M NL

1 of 1
AD A
093710

END
DATE
FILMED
2-81
DTIC



MICROCOPY RESOLUTION TEST CHART

ARO 16713.13-M

12

AD A093710

(6) SELECTING ALL TREATMENTS
BETTER THAN A CONTROL
USING EXISTING TABLES.

LEVEL II

(10) Roger L./Berger

(9) FSU Statistics Report M660
~~USARO~~ Technical Report, USARO D 49

(18) ARO

(11) October, 1980

(12) 27

Florida State University
Department of Statistics
Tallahassee, Florida 32306

(19) 16713.13-M

D-49-ARO

(14) FSU-STATISTICS-M660

(15) This research was supported in part by U.S. Army Research Office Grant DAAG29-79-C-0158.

81 1 12 043

400277

ABSTRACT

In this paper subset selection procedures for selecting all treatment populations with means larger than a control population are proposed. The treatments and control are assumed to have a multivariate normal distribution. Various covariance structures are considered. All of the proposed procedures are easily implemented using existing tables of the multivariate normal and multivariate t distributions. Some other procedures which have been proposed require extensive and unavailable tables for their implementation.

Key words: Multivariate normal, multivariate t, repeated measures, P*-condition.

SELECTING ALL TREATMENTS BETTER THAN
A CONTROL USING EXISTING TABLES

1. INTRODUCTION

Let $\Pi_1, \ldots, \Pi_k$ denote k ($k \geq 1$) treatment populations with means $\mu_1, \ldots, \mu_k$ and let Π_0 denote a control population with mean μ_0. It will be assumed that $\Pi_0, \ldots, \Pi_k$ have a multivariate normal distribution. Treatment population Π_i is said to be better than the control if $\mu_i \geq \mu_0$. The goal is to select a subset of the treatment populations which contains all populations which are better than the control. A correct selection (CS) is the selection of any subset which contains all the treatments which are better than the control. In this paper, selection procedures are proposed which insure that the probability of a correct selection, $P_{\underline{\mu}}(CS)$, is at least P^*, regardless of the true value of $\underline{\mu} = (\mu_0, \ldots, \mu_k)$, where P^* is a preassigned constant satisfying $0 < P^* < 1$. The requirement that $P_{\underline{\mu}}(CS) \geq P^*$ for all $\underline{\mu}$ is called the P^*-condition. The procedures proposed in this paper are easily implemented since any critical values needed can be obtained from existing tables of the multivariate normal distribution (e.g., Gupta, Nagel, and Panchapakesan (1973)) and multivariate t distribution (e.g., Krishnaiah and Armitage (1966)).

Paulson (1952) and Dunnett (1955) were among the first authors to consider treatment versus control comparison problems. Gupta and Sobel (1958) introduced the subset selection formulation which is being considered herein. Recently, Chen (1980) and Chen and Pickett (1980) have considered the subset selection formulation for the case of dependent populations. These authors have pointed out the importance of dependence in repeated measures designs.

This work is closely related to Chen (1980). It differs from Chen's in that some covariance structures are considered which Chen did not consider. In particular, this paper considers situations in which the control variance differs from the treatment variance. The selection procedures in this paper are the same as the procedures proposed by Chen in those situations when the same model is being considered. But the procedures are written in a slightly different form. This modified form has the advantage that existing tables for the multivariate normal and t distributions can now be used to implement the procedures. Thus, this work, in addition to proposing new selection procedures, should make some of Chen's procedures much easier to use. Existing tables can be used to implement the procedures for a wider range of models than the range of models for which tables were provided by Chen.

The following notation will be used. $\underline{Y} \sim MN(m, \underline{\mu}, \Sigma)$ means the random vector $\underline{Y}$ has an m-dimensional multivariate normal distribution with mean vector $\underline{\mu}$ and covariance matrix Σ. $\Phi(z)$ and $\phi(z)$ denote the distribution and density function of the standard univariate normal distribution. $\Phi_k(z_1, \ldots, z_k)$ denotes the distribution function of the k-variate standard normal distribution with zero means, unit variances and all correlations equal to ρ. $F_{k,\nu}(t_1, \ldots, t_k; \rho)$ denotes the distribution function of the k-variate central t distribution with ν degrees of freedom and all correlations equal to ρ.

2. KNOWN COVARIANCE CASE

In this section assume $\underline{X} \sim MN(k + 1, \underline{\mu}, V)$ where $\underline{X} = (X_0,\ldots,X_k)$; $\underline{\mu} = (\mu_0,\ldots,\mu_k)$ is unknown but $V = (v_{ij};\ i, j = 0,\ldots,k)$ is known. Further assume V has the form $v_{00} = v_0^2$, $v_{11} = \cdots = v_{kk} = v^2$, $v_{01} = \cdots = v_{0k} = a$, and $v_{ij} = b$ for $i \neq j$, $i,j = 1,\ldots,k$. Typically X_i, the observation from Π_i, will be a sample mean as the examples at this section's end illustrate but for now only the single vector observation $\underline{X}$ is considered. The k treatment populations are all assumed to have equal variances and covariances but the variance of the control, v_0^2, may be different and the covariance between the control and a treatment, a, need not equal the covariance between two treatments, b. Chen (1980) only considered the case in which $v_0^2 = v^2$. But in some situations much more data is available on the control than on the treatments. In these situations, it will usually be the case that $v_0^2 < v^2$.

2.1 Selection Procedure

Procedure R_1: Include population Π_i in the selected subset if and only if

$$x_i \geq x_0 - c_1\sqrt{v_0^2 + v^2 - 2a} \qquad (2.1)$$

where c_1 is chosen to satisfy (2.2).

Theorem 1: For a given P*, if c_1 is chosen to satisfy

$$\Phi_k(c_1,\ldots,c_1;\ \rho) = P^* \qquad (2.2)$$

where $\rho = (v_0^2 + b - 2a)/(v_0^2 + v^2 - 2a)$, then R_1 satisfies the P*-condition.

Proof: This proof is similar to the proof of Theorem 1 in Chen (1980). It is included here for completeness.

Fix $\underline{\mu} = (\mu_0,\ldots,\mu_k)$. Let $i_1,\ldots,i_B$ denote the subscripts of the B populations which are better than the control. Let $Z_i = (X_0 - X_i - (\mu_0 - \mu_i))/\sqrt{v_0^2 + v^2 - 2a}$, $i = 1,\ldots,k$. Then

$$\begin{aligned}
P_{\underline{\mu}}(CS|R_1) &= P_{\underline{\mu}}(\text{select } \Pi_{i_j},\ j = 1,\ldots,B) \\
&= P_{\underline{\mu}}(X_{i_j} \ge X_0 - c_1\sqrt{v_0^2 + v^2 - 2a}\ ,\ j = 1,\ldots,B) \\
&= P_{\underline{\mu}}(Z_{i_j} \le c_1 + (\mu_{i_j} - \mu_0)/\sqrt{v_0^2 + v^2 - 2a},\ j = 1,\ldots,B) \\
&\ge P_{\underline{\mu}}(Z_{i_j} \le c_1,\ j = 1,\ldots,B) \\
&\ge P_{\underline{\mu}}(Z_i \le c_1,\ i = 1,\ldots,k).
\end{aligned}$$

The first inequality is true since $\mu_{i_j} \ge \mu_0$, $j = 1,\ldots,B$. $\underline{Z} = (Z_1,\ldots,Z_k) \sim MN(k,\underline{0}, R)$ where $R = (r_{ij})$, $r_{ii} = 1$, $i = 1,\ldots, k$ and $r_{ij} = \rho$, $i \ne j$, $i,j = 1,\ldots,k$. By (2.2), $P_{\underline{\mu}}(Z_i \le c_1,\ i = 1,\ldots,k) = P^*$. Since $\underline{\mu}$ was arbitrary, R_1 satisfies the P^*-condition.||

2.2 Tables for c_1

The constant c_1 which depends on k, P*, and ρ is the value which is tabulated in Table 1 of Gupta, Nagel and Panchapakesan (1973). The correspondence of notation is $N = k$, $\alpha = 1 - P^*$ and $\rho = \rho$ where the notation on the left of each equality is the Gupta, Nagel and Panchapakesan notation and the notation on the right of each equality is the notation of this paper. This table covers P* = .75, .90, .975 and .99, all k values between 1 and 10 and all even k values between 12 and 50, and 17 different ρ values between .1 and .9. This table and interpolation therein seem to be adequate for the k and ρ values used in most applications. If other P* values are used, Table II of Gupta (1963) can be used. Here the correspondence of notation is $H = c_1$, $N = k$, $\rho = \rho$ and the tabled value is P* where again the left side of each equality is Gupta's notation and the right side is the notation of this paper.

If the value of c_1 for other values of k, P* and ρ is needed, then c_1 can be found by numerical methods as the solution of the equality

$$\int_{-\infty}^{\infty} \Phi^k((x\sqrt{\rho} + c_1)/\sqrt{1-\rho})\ d\Phi(x) = P^* \tag{2.3}$$

(see Gupta, Nagel and Panchapakesan (1973)). Solving (2.3) should be more efficient than solving equation (3.3) of Chen (1980) since (2.3) involves only a single integral whereas Chen's equation involves a double integral.

2.3 Examples

In the following examples, some special cases of the general model are considered. These examples illustrate some of the situations to which the general model applies. It should be remembered that in all these examples the procedures can be implemented easily since the constant c_1 can be obtained from Table 1 of Gupta, Nagel and Panchapakesan (1973).

Example 1: Let $\underline{Y}_1,\ldots,\underline{Y}_n$ be independent. $Y_i \sim MN(k+1, \underline{\mu}, \Sigma)$ where $\Sigma = (\sigma_{ij};\ i, j = 0,\ldots,k)$. Further assume Σ has the form $\sigma_{00} = \sigma_0^2$, $\sigma_{11} = \ldots = \sigma_{kk} = \sigma^2$, $\sigma_{01} = \ldots = \sigma_{0k} = \alpha$ and $\sigma_{ij} = \beta$, for $i \neq j$, $i, j = 1,\ldots,k$. Let $\underline{X}$ be the sample mean of $\underline{Y}_1,\ldots,\underline{Y}_n$. Then $\underline{X} \sim MN(k+1, \underline{\mu}, V)$ where $v_0^2 = \sigma_0^2/n$, $v^2 = \sigma^2/n$, $a = \alpha/n$ and $b = \beta/n$. The procedure R_1 becomes select Π_i if and only if

$$x_i \geq x_0 - c_1\sqrt{(\sigma_0^2 + \sigma^2 - 2\alpha)/n} \tag{2.4}$$

where c_1 is determined by (2.2) with $\rho = (\sigma_0^2 + \beta - 2\alpha)/(\sigma_0^2 + \sigma^2 - 2\alpha)$.

The case in which $\sigma_0 = \alpha = 0$ is of special interest. In this case, X_0 equals μ_0 with probability one. That is to say, this is the case in which the control mean μ_0 is known. In this case, R_1 is select Π_i if and only if

$$x_i \geq \mu_0 - c_1\sigma/\sqrt{n} \tag{2.5}$$

where c_1 is determined by (2.2) with $\rho = \beta/\sigma^2$, the correlation between any two treatment populations. This procedure is the procedure P_1 proposed by Chen (1980) for the μ_0 known case.

<u>Example 2</u>: Assume the same model as in Example 1. Assume further that $\sigma_0 = \sigma$ and $\alpha = \beta$. Table I in Chen (1980) was provided for this equal variance and equal covariance case. Let $\gamma = \alpha/\sigma^2$ be the common known correlation. The procedure R_1 becomes select Π_i if and only if

$$x_i \geq x_0 - c_1\sigma\sqrt{2(1 - \gamma)/n} \tag{2.6}$$

where c_1 is determined by (2.2) with $\rho = 1/2$. Comparing R_1 with the procedure P_2 proposed by Chen (1980) for this case, they are found to be the same when the identification $d_2 = c_1\sqrt{2(1 - \gamma)}$ is made. d_2 is the constant tabled by Chen. The advantage of writing the procedure in the form (2.6) is that, whereas Chen required a separate table entry for each value of γ(ρ in Chen's notation), only the Gupta, Nagel and Panchapakesan (1973) table for $\rho = 1/2$ is needed to determine c_1, regardless of the value of γ. The form (2.6) and the Gupta, Nagel and Panchapakesan table might also be preferred since this table provides four decimal places for c_1 whereas Chen's Table I provides only two decimal places for d_2.

<u>Example 3</u>: Procedure R_1 can be used in the situation in which there are separate samples of different sizes on the control and treatment populations. It can be used if in addition there is a joint sample on the control and treatment populations. Let $\underline{Y}_1,\ldots,\underline{Y}_n$ be defined as in Example 1. Let m_1, m_2 and m_3 be non-negative integers with $m_1 + m_2 + m_3 = n$. Let $r = m_1 + m_2$ and $s = m_1 + m_3$. Let $X_0 = \sum_{j=1}^{r} Y_{0j}/r$ and $X_i = (\sum_{j=1}^{m_1} Y_{ij} + \sum_{j=r+1}^{n} Y_{ij})/s$, $i = 1,\ldots,k$.

The sample size for the joint sample of the treatments and the control is m_1. The sample sizes of the additional samples on the control and the treatments are m_2 and m_3 respectively. Then $\underline{X} \sim MN(k + 1, \underline{\mu}, V)$ where $v_0^2 = \sigma_0^2/r$, $v^2 = \sigma^2/s$, $a = m_1\alpha/rs$ and $b = \beta/s$. For this model, R_1 becomes select Π_i if and only if

$$x_i \geq x_0 - c_1 \sqrt{(s\sigma_0^2 + r\sigma^2 - 2m_1\alpha)/rs} \tag{2.7}$$

where c_1 is determined by (2.2) with $\rho = (s\sigma_0^2 + r\beta - 2m_1\alpha)/(s\sigma_0^2 + r\sigma^2 - 2m_1\alpha)$.

A case of particular interest is the case $m_1 = 0$. This is the case in which there is a sample of size m_2 from the control population and an independent sample of size m_3 from the treatment populations. If in addition the treatment populations are independent, then R_1 reduces to the procedure proposed by Gupta and Sobel (1958) (equation (3.10)) if the identification is made that $d = c_1\sqrt{m_3\sigma_0^2 + m_2\sigma^2}/\sigma\sqrt{m_2}$ where d is a constant defined by Gupta and Sobel. The Gupta and Sobel procedure may be used when there are unequal sample sizes on the various treatments, a situation not covered by the model presented here.

3. UNKNOWN VARIANCE, KNOWN CORRELATION CASE

In this section the case in which the treatments and control have a common unknown variance and known correlations is considered.

Let $\underline{Y}_1,\ldots,\underline{Y}_n$ be independent. $\underline{Y}_i \sim MN(k+1, \underline{\mu}, \sigma^2 R)$ where $\underline{\mu} = (\mu_0,\ldots,\mu_k)$ and σ^2 are unknown but $R = (r_{ij};\ i, j = 0,\ldots,k)$ is known and has the form $r_{00} = \ldots = r_{kk} = 1$, $r_{01} = \ldots = r_{0k} = r_0$ and $r_{ij} = r$ for $i \neq j$, $i, j = 1,\ldots,k$. Let $\underline{X} = (X_0,\ldots,X_k)$ be the sample mean of $\underline{Y}_1,\ldots,\underline{Y}_n$. Let $S = (s_{ij};\ i, j = 0,\ldots,k)$ be the usual unbiased sample covariance matrix, i.e., $s_{ij} = \sum_{m=1}^{n} (y_{im} - x_i)(y_{jm} - x_j)/(n-1)$, $i, j = 0,\ldots,k$. An estimate of σ^2 which will be used is $S_0^2 = tr(R^{-1}S)/(k+1)$. It is known (see Anderson (1958)) that S_0^2 is independent of $\underline{X}$ and $(k+1)(n-1)S_0^2/\sigma^2$ has a chi-squared distribution with $\nu = (k+1)(n-1)$ degrees of freedom. For computational purposes, it should be noted that

$$tr(R^{-1}S) = \frac{ds_{00} + 2e(\sum_{i=1}^{k} s_{i0}) + f(\sum_{i=1}^{k} s_{ii}) + 2g(\sum_{i>j\geq 1} s_{ij})}{1 + (k-1)r - kr_0^2}$$

where $d = 1 + (k-1)r$, $e = -r_0$, $g = (r_0^2 - r)/(1-r)$ and

$f = 1 + (k - 1 - r_0^2)r - (k-1)r_0^2 - (k-1)r(r_0^2 - r)/(1-r)$.

3.1 Selection Procedure

Procedure R_2: Include population Π_i in the selected subset if and only if

$$x_i \geq x_0 - c_2 s_0 \sqrt{(2 - 2r_0)/n} \tag{3.1}$$

where c_2 is chosen to satisfy (3.2).

<u>Theorem 2</u>: <u>For a given</u> P^*, <u>if</u> c_2 <u>is chosen to satisfy</u>

$$F_{k,\nu}(c_2,\ldots,c_2;\ \rho) = P^* \qquad (3.2)$$

<u>where</u> $\rho = (1 + r - 2r_0)/(2 - 2r_0)$ <u>and</u> $\nu = (k + 1)(n - 1)$, <u>then</u> R_2 <u>satisfies the</u> P^*<u>-condition</u>.

<u>Proof</u>: Fix $\underline{\mu} = (\mu_0,\ldots,\mu_k)$ and σ^2. Let

$Z_i = (X_0 - X_i - (\mu_0 - \mu_i))/\sqrt{(2 - 2r_0)/n}$, $i = 1,\ldots,k$. Let $T_i = Z_i/S_0$. Then $\underline{Z} = (Z_1,\ldots,Z_k) \sim MN(k,\underline{0},V)$ where $V = (v_{ij};i,\ j = 1,\ldots,k)$, $v_{ii} = \sigma^2$, $i = 1,\ldots,k$ and $v_{ij} = \sigma^2\rho$, $i \neq j$, $i,\ j = 1,\ldots,k$, and $(k + 1)(n - 1)\ S_0^2/\sigma^2$ has a chi-squared distribution with ν degrees of freedom and is independent of $\underline{Z}$. Thus $\underline{T} = (T_1,\ldots,T_k)$ has a standard central multivariate t distribution with ν degrees of freedom and all the off diagonal elements of the correlation matrix equal to ρ. Arguing as in the proof of Theorem 1,

$$\begin{aligned} P_{\underline{\mu},\sigma}(CS|R_2) &\geq P_{\underline{\mu},\sigma}(T_i \leq c_2,\ i = 1,\ldots,k) \\ &= F_{k,\nu}(c_2,\ldots,c_2;\ \rho) = P^*. \end{aligned}$$

Since $\underline{\mu}$ and σ^2 were arbitrary, R_2 satisfies the P^*-condition.||

<u>3.2 Tables for c_2</u>

The constant c_2 which depends on k, P^*, ρ and ν is the value which is tabulated in Krishnaiah and Armitage (1966). The correspondence of notation is $p = k$, $\alpha = 1 - P^*$, $\rho = \rho$ and $n = \nu$ where the notation on the left of each equality is the Krishnaiah and Armitage notation and the notation on the right of each equality is the notation of this paper. This table covers $P^* = .95$ and $.99$, $k = 1(1)10$, $\rho = 0.0(.1).9$ and $\nu = 5(1)35$. For larger values of ν, Table I of Gupta, Nagel and Panchapakesan (1973) may be used to approximate c_2 since this normal table corresponds to $\nu = \infty$ (cf. Section 2.2). Gupta (1963a) provides references to some other partial tables of the multivariate t distribution.

If the value of c_2 for other values of k, P*, ρ and ν is needed, it can be found by numerical methods as the solution of the equality

$$\int_0^\infty h_\nu(\chi) \int_{-\infty}^{\infty} \Phi^k((x\sqrt{\rho} + c_2\chi/\sqrt{\nu})/\sqrt{1-\rho})\phi(x)\, dx\, d\chi = P^* \tag{3.3}$$

where $h_\nu(\chi)$ is the chi density corresponding to ν degrees of freedom for the chi-squared distribution (see equation (6.7) of Gupta (1963b)). Solving (3.3) should be more efficient than solving equation (5.3) of Chen (1980) since (3.3) involves only a double integral whereas Chen's equation involves a triple integral.

3.3 Example

Example 4: Procedure R_2 is the same as the procedure P_4 proposed by Chen (1980) if the identification is made that $d_4 = c_2\sqrt{2 - 2r_0}$ where d_4 is a constant defined by Chen. Chen's procedure was proposed for a more general correlation structure. But the advantage of writing the procedure as R_2 is that c_2 depends only on r_0 and r through ρ whereas a separate value of d_4 is required for each r_0 and r pair. In particular, assume $r_0 = r$. Then $\rho = 1/2$. This is the case for which Table II of Chen is provided. Whereas Table II requires a separate entry for each value of r (ρ in Chen's notation), only the Krishnaiah and Armitage (1966) table for $\rho = 1/2$ is needed when procedure R_2 is used. The Krishnaiah and Armitage table also provides percentage points for many more values of ν than does Table II.

4. UNKNOWN VARIANCE, UNKNOWN CORRELATION CASE

In this section the case is considered in which the treatments and control have a common unknown variance and a common unknown correlation.

Let $\underline{Y}_1,\ldots,\underline{Y}_n$ be defined as in Section 3. Assume $r = r_0$, that is, the correlation between a treatment and the control is equal to the correlation between two treatments. But now assume r is unknown. This model might be used in a repeated measures design in which each of the k + 1 observations in $\underline{Y}_i$ are observations on the same individual or experimental unit. Let $\underline{X}$ be the sample mean of $\underline{Y}_1,\ldots,\underline{Y}_n$. Each of the variables $X_0 - X_i$, $i = 1,\ldots,k$, has the variance $2\sigma^2(1 - r)/n$. Let $S_1^2 = \sum_{i=1}^{n} (Y_{0i} - Y_{1i} - (X_0 - X_1))^2/(n - 1)$. S_1^2 will be used as an estimate of $2\sigma^2(1 - r)$.

4.1 Selection Procedure

Procedure R_3: Include population Π_i in the selected subset if and only if

$$x_i \geq x_0 - c_3 s_1/\sqrt{n} \tag{4.1}$$

where c_3 is chosen to satisfy (4.2).

Theorem 3: For a given P*, if c_3 is chosen to satisfy

$$F_{k,\, n-1}(c_3,\ldots,c_3;\, 1/2) = P^*, \tag{4.2}$$

then R_3 satisfies the P*-condition.

Proof: Let $\underline{U}_1,\ldots,\underline{U}_k$ be defined by $U_{ij} = Y_{0j} - Y_{ij}$, $i = 1,\ldots,k$; $j = 1,\ldots,n$. Let $\underline{W}$ be the sample mean of $\underline{U}_1,\ldots,\underline{U}_n$. Then $W_i = X_0 - X_i$. S_1^2 is the upper left corner element of the sample covariance matrix computed from $\underline{U}_1,\ldots,\underline{U}_n$. Thus S_1^2 is independent of $\underline{W}$. The elements of $\underline{W}$ are equally correlated with correlation equal to 1/2. Using these facts, the proof is now similar to the proof of Theorem 2. ||

The constant c_3 can be obtained from the table of Krishnaiah and Armitage (1966) as explained in Section 3.2. Only the $\rho = 1/2$ table is needed to obtain c_3. The table of Gupta and Sobel (1957) can also be used to obtain c_3. The correspondence of notation is $p = k$, $P^* = P^*$, $\nu = \nu$ and $q/\sqrt{2} = c_3$ where the notation on the left of each equality is the Gupta and Sobel notation and the notation on the right is the notation of this paper. This table covers $P^* = .75$, $.90$ and $.975$, values not covered by the Krishnaiah and Armitage table.

The use of S_1^2 as an estimate of $2\sigma^2(1 - r)$ is not entirely satisfactory. Any of the statistics $S_j^2 = \sum_{i=1}^{n} (Y_{0i} - Y_{ji} - (X_0 - X_j))^2/(n - 1)$, $j = 1,\ldots,k$ could be used. S_1^2 was chosen arbitrarily. It would be good to combine the S_j^2's to get a better estimate. But the S_j^2's are not independent so their sum may not have a chi-squared distribution. If n is large then $S^2 = \sum_{j=1}^{k} S_j^2/k$ may be used in place of S_1^2 in procedure R_3 and c_3 may be approximated by the value in Table 1 of Gupta, Nagel and Panchapakesan (1973). This is valid since S^2 converges to $2\sigma^2(1 - r)$ in probability as $n \to \infty$.

5. FURTHER COMMENTS

Each of the procedures R_1, R_2 and R_3 have this form. Include population Π_i in the selected subset if and only if

$$x_i \geq x_0 - cSE(X_0 - X_i)$$

where c is an appropriate constant and $SE(X_0 - X_i)$ is the standard deviation of $X_0 - X_i$ or an estimate thereof. This form reduces the number of parameters upon which the constant c depends. For example, in Section 2 the constant c does not depend on the parameter γ whereas, if the rule is written in the form of Chen (1980), the constant does depend on γ. Berger and Gupta (1980) found that the use of the standard deviation of the differences, $X_0 - X_i$, had other advantages in a different subset selection problem. This consideration of the differences as the important variables and use of their standard deviations may be advantageous in other similar problems.

BIBLIOGRAPHY

Anderson, T.W. (1958). An Introduction to Multivariate Statistical Analysis. New York: John Wiley & Sons, Inc.

Berger, R.L. and Gupta, S.S. (1980). Minimax subset selection rules with applications to unequal variance (unequal sample size) problems. Scand. J. Statist., 7, 21-26.

Chen, H.J. (1980). On selecting a subset which contains all populations better than a control. Comm. Statist. A Theory Methods, A9, 851-864.

Chen, H.J. and Pickett, J.R. (1980). Selecting a subset which contains all populations better than a control in repeated measurements designs. Technical report.

Dunnett, C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. J. Amer. Statist. Assoc., 50, 1096-1121.

Gupta, S.S. (1963a). Bibliography of the multivariate normal integrals and related topics. Ann. Math. Statist., 34, 829-838.

_____ (1963b). Probability integrals of multivariate normal and multivariate t. Ann. Math. Statist., 34, 792-828.

Gupta, S.S., Nagel, K. and Panchapakesan, S. (1973). On the order statistics from equally correlated normal random variables. Biometrika, 60, 403-413.

Gupta, S.S. and Sobel, M. (1957). On a statistic which arises in selection and ranking problems. Ann. Math. Statist., 28, 957-967.

Gupta, S.S. and Sobel, M. (1958). On selecting a subset which contains all populations better than a standard. Ann. Math. Statist., 29, 235-244.

Krishnaiah, P.R. and Armitage, J.V. (1966). Tables for multivariate t distribution. Sankhya B, 28, 31-56.

Paulson, E. (1952). On the comparison of several experimental categories with a control. Ann. Math. Statist., 23, 610-616.

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER FSU No. M660 USARO No. D-49	2. GOVT ACCESSION NO. AD-A093 710	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and subtitle) Selecting All Treatments Better Than a Control Using Existing Tables		5. TYPE OF REPORT & PERIOD COVERED Technical report 6. PERFORMING ORG. REPORT NUMBER FSU Statistics Report M660
7. AUTHOR(s) Roger L. Berger		8. CONTRACT OR GRANT NUMBER(s) DAAG29 79 C 0158
9. PERFORMING ORGANIZATION NAME AND ADDRESS Florida State University Department of Statistics Tallahassee, FL. 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office P.O. Box 12211 Research Triangle Park, N.C. 27709		12. REPORT DATE October, 1980 13. NUMBER OF PAGES 14
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS

Multivariate normal, multivariate t, repeated measures, P*-condition

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper subset selection procedures for selecting all treatment populations with means larger than a control population are proposed. The treatments and control are assumed to have a multivariate normal distribution. Various covariance structures are considered. All of the proposed procedures are easily implemented using existing tables of the multivariate normal and multivariate t distributions. Some other procedures which have been proposed require extensive and unavailable tables for their implementation.